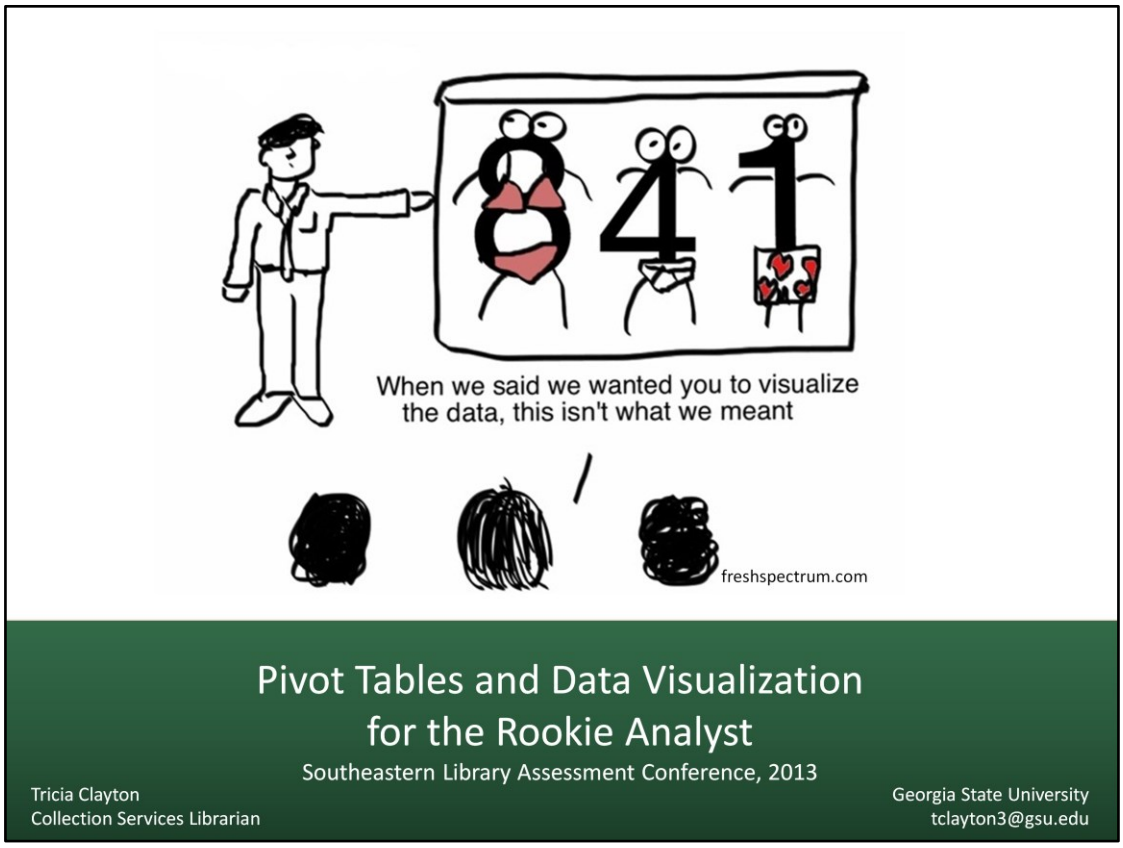


Pivot Tables and Data Visualization for the Rookie Analyst

Southeastern Library Assessment Conference, 2013

Tricia Clayton
Collection Services Librarian

Georgia State University
tclayton3@gsu.edu

Presented at the Southeastern Library Assessment Conference, October 22, 2013

Agenda

- Pivot tables
  - What are they?
  - Creating them
  - Data manipulation & analysis
- Data visualization
  - Selected best practices
  - Choosing appropriate charts and graphs

Image source: http://en.mycartoons.de/cartoon/to-do-list

Pivot tables

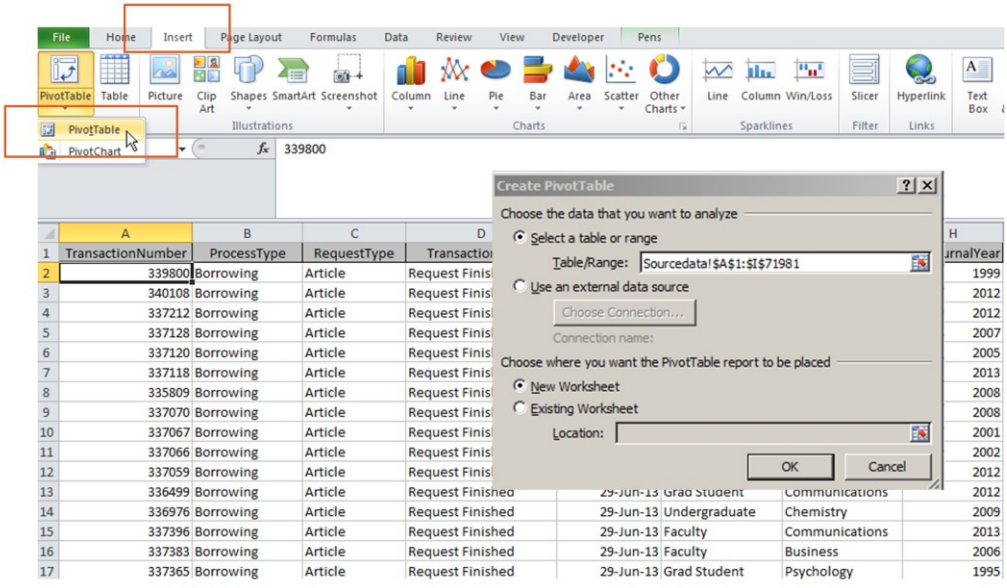
- Quick and easy reporting tool to summarize and analyze your data
- Great for large data sets
- Dynamic – add, delete, and rearrange (pivot) columns and rows without altering your source data

ILL transactions FY12 and FY13

	A	B	C	D	E	F	G	H
1	TransactionNumber	ProcessType	RequestType	TransactionStatus	TransactionDate	Status	Department	JournalYear
2	339800	Borrowing	Article	Request Finished	29-Jun-13	Grad Student	Health Sciences	1999
3	340108	Borrowing	Article	Request Finished	29-Jun-13	Faculty	Psychology	2012
4	337212	Borrowing	Article	Request Finished	29-Jun-13	Faculty	Criminal Justice	2012
5	337128	Borrowing	Article	Request Finished	29-Jun-13	Grad Student	Education	2007
6	337120	Borrowing	Article	Request Finished	29-Jun-13	Grad Student	Education	2005
7	337118	Borrowing	Article	Request Finished	29-Jun-13	Grad Student	Education	2013
8	335809	Borrowing	Article	Request Finished	29-Jun-13	Grad Student	Psychology	2008
9	337070	Borrowing	Article	Request Finished	29-Jun-13	Grad Student	Education	2008
71929	262741	Borrowing	Article	Cancelled by ILL Staff	01-Jul-11	Faculty	Other	1988
71930	262765	Lending	Loan	Cancelled by ILL Staff	01-Jul-11	None	Non GSU	
71931	262763	Lending	Article	Cancelled by ILL Staff	01-Jul-11	None	Non GSU	2011
71932	262706	Lending	Loan	Cancelled by ILL Staff	01-Jul-11	None	Non GSU	
71933	262702	Lending	Loan	Cancelled by ILL Staff	01-Jul-11	None	Non GSU	
71934	259560	Borrowing	Loan	Request Finished	01-Jul-11	Faculty	Communications	
71935	257245	Borrowing	Loan	Request Finished	01-Jul-11	Grad Student	Biology	
71936	259139	Borrowing	Loan	Request Finished	01-Jul-11	Staff	Biology	
71937	258798	Borrowing	Loan	Request Finished	01-Jul-11	Grad Student	Education	
71938	260338	Borrowing	Loan	Request Finished	01-Jul-11	Undergraduate	Other	
71939	260437	Borrowing	Loan	Request Finished	01-Jul-11	Faculty	History	
71940	257984	Borrowing	Loan	Request Finished	01-Jul-11	Faculty	Music	
71941	258800	Borrowing	Loan	Request Finished	01-Jul-11	Grad Student	Education	
71942	262313	Borrowing	Loan	Request Finished	01-Jul-11	Undergraduate	History	2004
71943	260164	Borrowing	Loan	Request Finished	01-Jul-11	Faculty	Business	
71944	260165	Borrowing	Loan	Request Finished	01-Jul-11	Faculty	Business	

This file contains 2 years of our interlibrary loan transactions downloaded from ILLiad.
70,000+ rows, multiple fields = an ideal file for pivot tables

Creating a pivot table

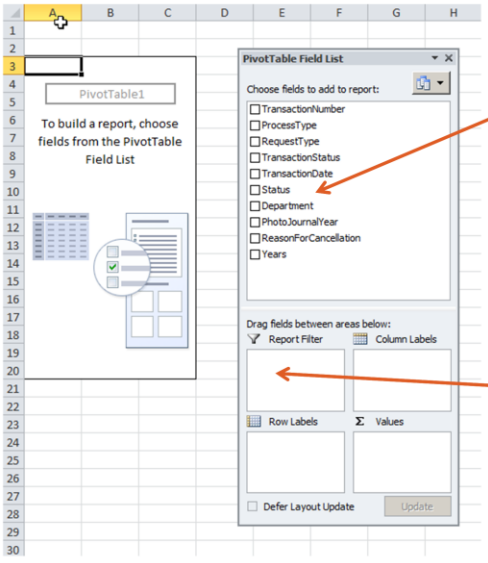


	A	B	C	D				H
1	TransactionNumber	ProcessType	RequestType	Transactio				urnalYear
2	339800	Borrowing	Article	Request Finis				1999
3	340108	Borrowing	Article	Request Finis				2012
4	337212	Borrowing	Article	Request Finis				2012
5	337128	Borrowing	Article	Request Finis				2007
6	337120	Borrowing	Article	Request Finis				2005
7	337118	Borrowing	Article	Request Finis				2013
8	335809	Borrowing	Article	Request Finis				2008
9	337070	Borrowing	Article	Request Finis				2008
10	337067	Borrowing	Article	Request Finis				2001
11	337066	Borrowing	Article	Request Finis				2002
12	337059	Borrowing	Article	Request Finis				2012
13	336499	Borrowing	Article	Request Finished	29-Jun-13	Grad Student	Communications	2012
14	336976	Borrowing	Article	Request Finished	29-Jun-13	Undergraduate	Chemistry	2009
15	337396	Borrowing	Article	Request Finished	29-Jun-13	Faculty	Communications	2013
16	337383	Borrowing	Article	Request Finished	29-Jun-13	Faculty	Business	2006
17	337365	Borrowing	Article	Request Finished	29-Jun-13	Grad Student	Psychology	1995

From the ribbon – Insert – PivotTable

By default, Excel selects the full range of your data on the active worksheet, and places it in a new worksheet. You can edit this.

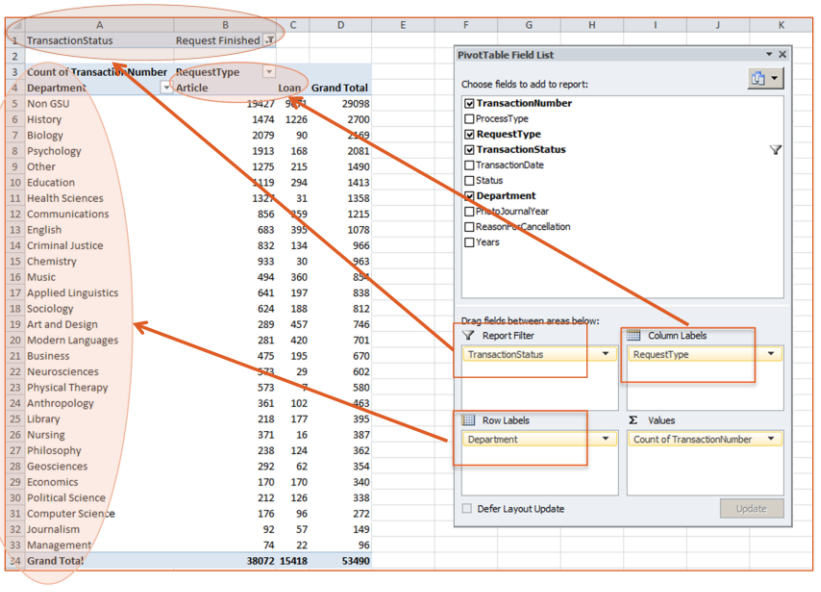
Empty pivot table



Select fields or drag them to one of the four drop zones below to add them to your table.

The **Report Filte**r drop zone is a top level filter for the entire chart. You can also filter elements in individual columns and rows.

Elements of a pivot table

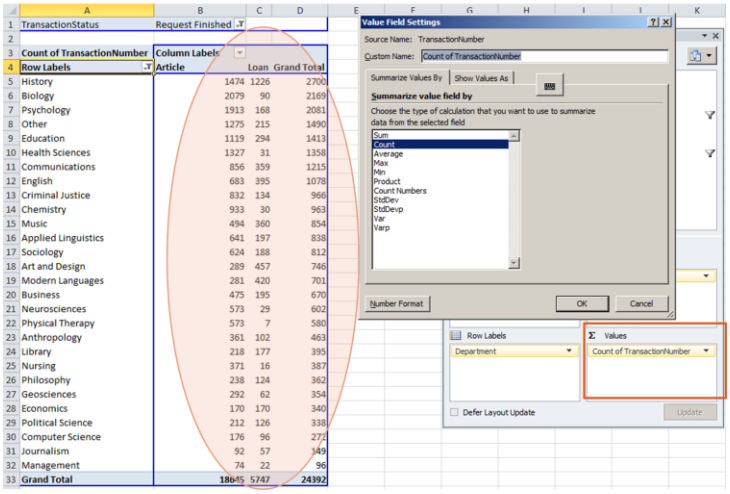


This is an example of a simple pivot table: the entire table is filtered by Transaction Status for finished requests; Requests Types (articles and loans) are in the columns, and Departments are in the rows.

You can add multiple row/column fields. Collapse/expand buttons appear if there is more than one field in the row/column drop zones.

Excel automatically adds filter arrows to the Report Filter field as well as to the column and row headings.

Value field settings



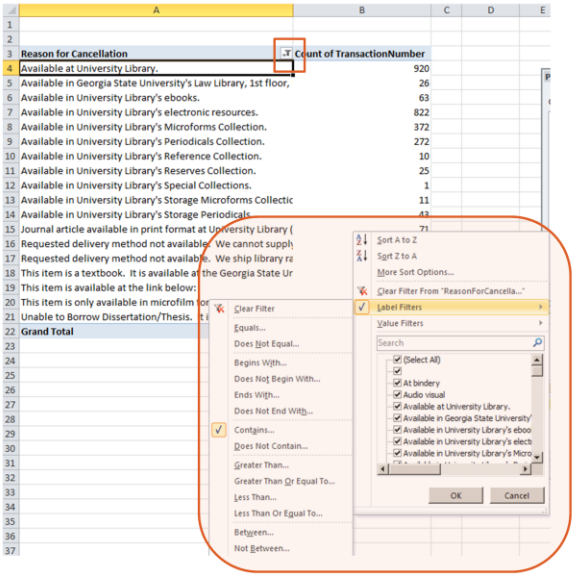
TransactionStatus	Request Finished
Count of TransactionNumber	Column Labels
Row Labels	Article	Loan	Grand Total
History	1474	1226	2700
Biology	2079	90	2169
Psychology	1913	168	2081
Other	1275	215	1490
Education	1119	294	1413
Health Sciences	1327	31	1358
Communications	856	359	1215
English	683	395	1078
Criminal Justice	832	134	966
Chemistry	933	30	963
Music	494	360	854
Applied Linguistics	641	197	838
Sociology	624	188	812
Art and Design	289	457	746
Modern Languages	281	420	701
Business	475	195	670
Neurosciences	573	29	602
Physical Therapy	573	7	580
Anthropology	361	102	463
Library	218	177	395
Nursing	371	16	387
Philosophy	238	124	362
Geosciences	292	62	354
Economics	170	170	340
Political Science	212	126	338
Computer Science	176	96	272
Journalism	92	57	149
Management	74	22	96
Grand Total	18615	5747	24392

The **Values** area determines what data is presented in the cells of the table and summarized in the last column (totaled by default).

Choices for summarizing your data (in the Value Field Settings) include: sum, count, minimum, maximum, and average.

In this context, **average** is the arithmetic mean - the sum of all the numbers in a series divided by the count of the numbers in that series.

Filtering, sorting, grouping…

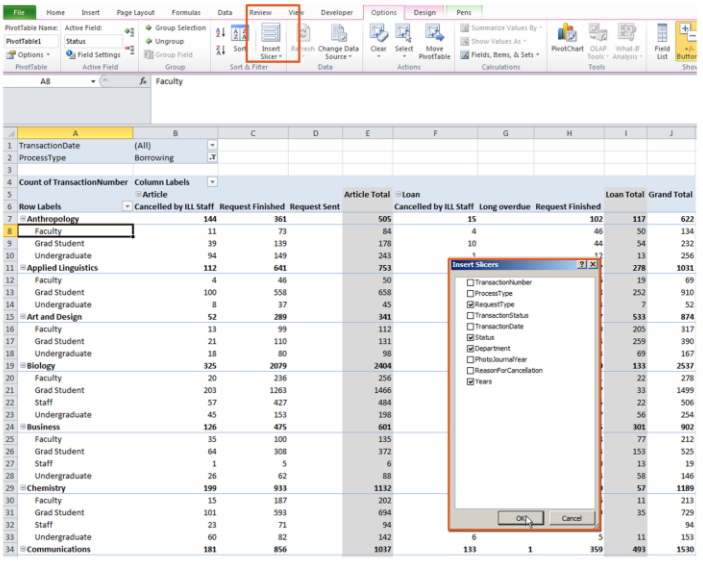


Sorting and more advanced filtering options are available from the filter icons or by right clicking in the field.

Filtering options are contextual, i.e. label filters for text fields and date filters for dates.

Right click in a field to see grouping options.

Slicers – easily filter multiple fields



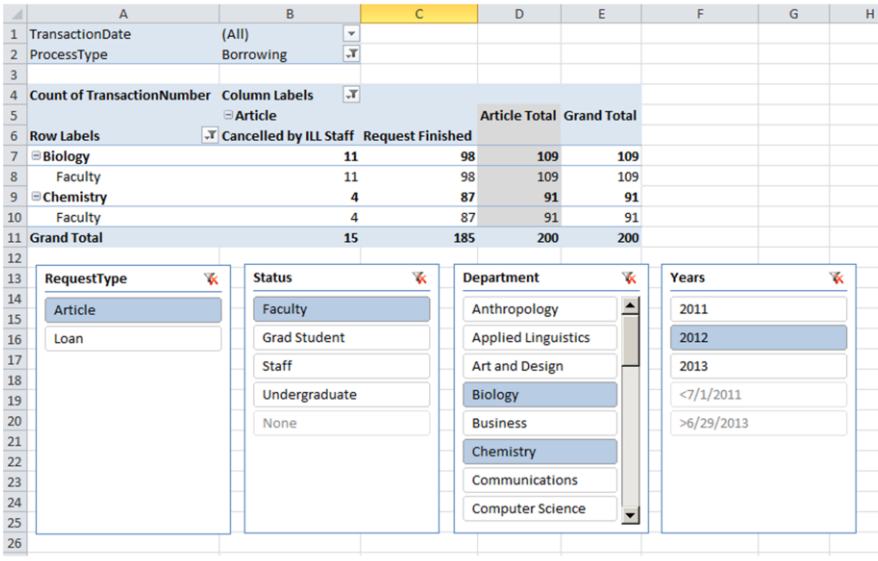
TransactionDate	(All)								
ProcessType	Borrowing								
Count of TransactionNumber	Column Labels								
	Article				Loan				
Row Labels	Cancelled by ILL Staff	Request Finished	Request Sent	Article Total	Cancelled by ILL Staff	Long overdue	Request Finished	Loan Total	Grand Total
Anthropology	144	361		505	15		102	117	622
Faculty	11	73		84	4		46	50	134
Grad Student	39	139		178	10		44	54	232
Undergraduate	94	149		243	1		12	13	256
Applied Linguistics	112	641		753				278	1031
Faculty	4	46		50				19	69
Grad Student	100	558		658				252	910
Undergraduate	8	37		45				7	52
Art and Design	52	289		341				533	874
Faculty	13	99		112				205	317
Grad Student	21	110		131				259	390
Undergraduate	18	80		98				69	167
Biology	325	2079		2404				133	2537
Faculty	20	236		256				22	278
Grad Student	203	1263		1466				33	1499
Staff	57	427		484				22	506
Undergraduate	45	153		198				56	254
Business	126	475		601				301	902
Faculty	35	100		135				77	212
Grad Student	64	308		372				153	525
Staff	1	5		6				13	19
Undergraduate	26	62		88				58	146
Chemistry	199	933		1132				57	1189
Faculty	15	187		202				11	213
Grad Student	101	593		694				35	729
Staff	23	71		94					94
Undergraduate	60	82		142	6		5	11	153
Communications	181	856		1037	133	1	359	493	1530

This pivot table still has a lot of data to digest – it counts borrowing requests by their status for both loans and articles (columns), listed by department then by patron status (rows). Slicers help to quickly zero in on specifics.

Slicers are a newer feature, introduced in Office 2010. Not available in the Mac (as of Office 2011).

From the ribbon – Insert – Slicers; then select desired fields.

Slicers filter four categories to count only: science faculty article requests in 2012



TransactionDate	(All)
ProcessType	Borrowing

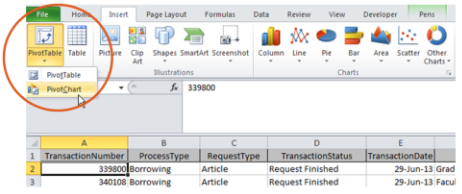
Count of TransactionNumber	Column Labels			
	Article		Article Total	Grand Total
Row Labels	Cancelled by ILL Staff	Request Finished		
Biology	**11**	**98**	**109**	**109**
Faculty	11	98	109	109
Chemistry	**4**	**87**	**91**	**91**
Faculty	4	87	91	91
Grand Total	**15**	**185**	**200**	**200**

RequestType	Status	Department	Years
Article	Faculty	Anthropology	2011
Loan	Grad Student	Applied Linguistics	2012
	Staff	Art and Design	2013
	Undergraduate	Biology	<7/1/2011
	None	Business	>6/29/2013
		Chemistry	
		Communications	
		Computer Science	

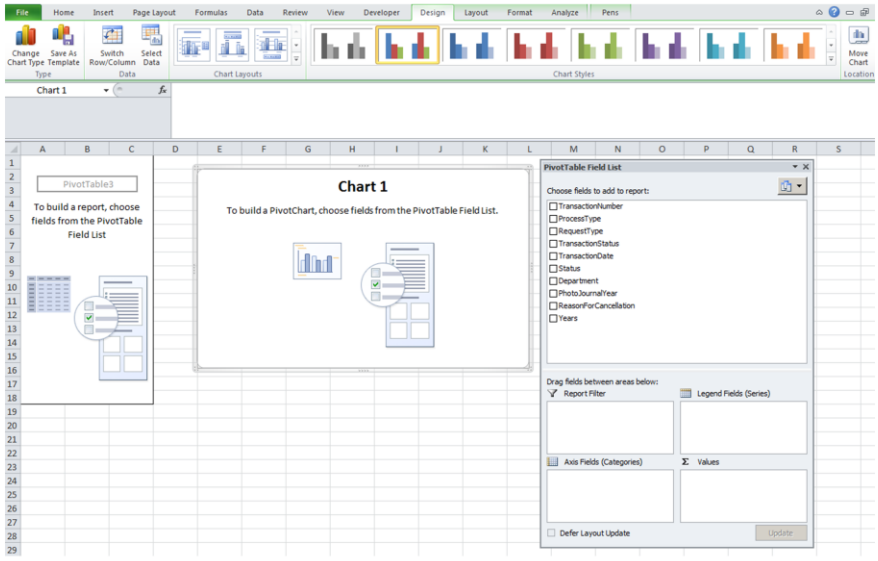
Pivot charts

- Pivot charts graphically display the summary data of pivot tables. Like the tables whose data they are displaying, they are dynamic.

- They can be created simultaneously with a pivot table, or afterwards by clicking inside the pivot table before creating the chart.

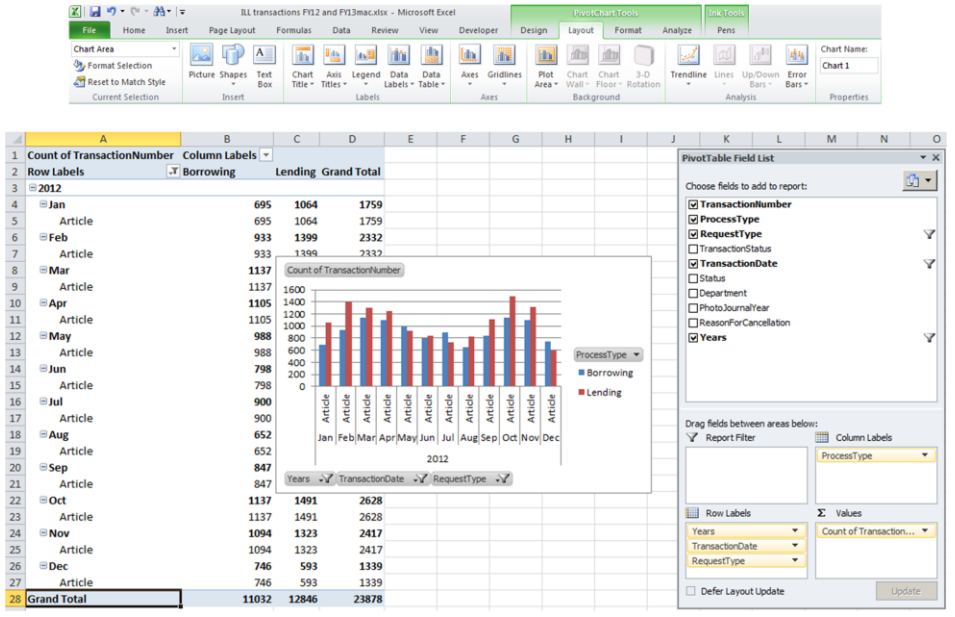


Empty pivot table/chart on a new worksheet



The initial layout looks similar to the one for an empty pivot table, except there is an added center placeholder for the chart. The chart design tools are active in the ribbon.

Borrowing and lending transactions in 2012



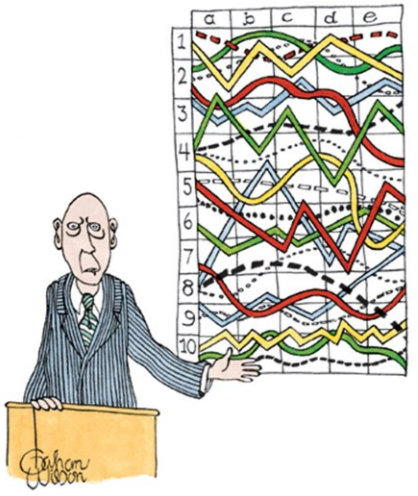
The default chart isn't very pretty, but it helps you quickly visualize your data.

Can filter fields directly on the chart (changes appear in both the chart and the table).

Right click on the chart to change chart types.

More options are available in the PivotChart Tools section of the ribbon for editing chart format, layout etc.

Data visualization



"I'll pause for a moment so you can let this information sink in."

Source: New Yorker, http://www.newyorker.com/humor/issuecartoons/2010/12/06/cartoons_20101129#slide=7

Definitions of data visualization vary throughout the literature – for this presentation I'm thinking in simple terms: the act of analyzing and/or presenting data visually

Visualization tools include:
-Excel (what I primarily use)
-Google Spreadsheets
-Many Eyes - http://www-958.ibm.com (IBM online project, encourages group thinking about visualizations)
-Tableau (Windows only, need the commercial version to keep your data private)
-programming solutions such as R (open source statistical programming software) and SAS ($)

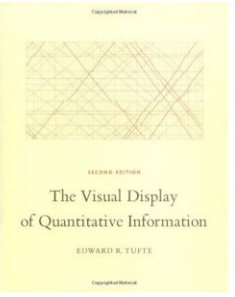
Today's scope is just the tip of the iceberg for this topic: to review a selection of best practices and some appropriate ways to showcase your data – not exhaustive!

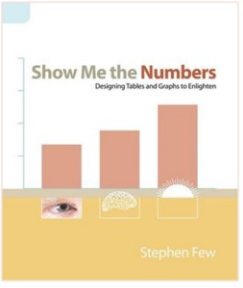
I'm not covering: mapping, info graphics, math/statistical topics - but basic best practices for design would still apply.

I'm not covering: considerations like "what question are you trying to answer?", "who is your audience?", "what story are you telling?" – though these are critical things to think about in your overall activities of creating visualizations for your administrators and stakeholders.

Data visualization: selected best practices

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.





Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten* (2nd ed.). Burlingame, CA: Analytics Press.

For recommended reading about data visualization, either or both of these are great starting points; they were my best sources for this presentation.

Edward Tufte – one of the most important thinkers in the modern field of data visualization. *The visual display of quantitative information* was originally published in 1986, and is a classic study on theory and practice in the design of data graphics.

Stephen Few – another leading expert in the field, 25+ years experience; educator and consultant, founder of the company Perceptual Edge.
His first book, *Show Me the Numbers: Designing Tables and Graphs to Enlighten* (2004; 2012) is a comprehensive and practical guide to business graphics.

Graphical Excellence

Excellence in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency.

Graphical displays should:

- show the data
- avoid distorting the data
- make large data sets coherent
- induce the viewer to think about the substance, not the methodology or design

(Tufte, 2001)

Tufte writes a lot about graphical excellence. At its core, "excellence begins with telling the truth about the data."

Graphical Excellence

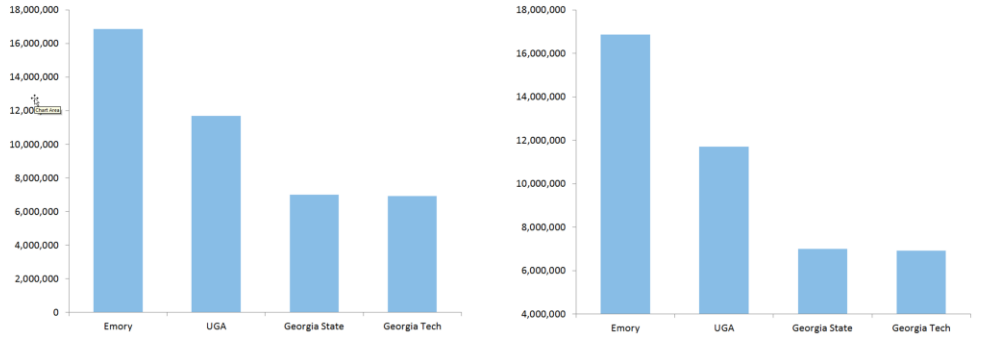
> The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.
>
> (Tufte, 2001)

$$\text{Lie Factor} = \frac{\text{size of effect shown in the graphic}}{\text{size of effect in the data}}$$

Tufte coined the term lie factor. A factor equal to one is ideal.

Graphical integrity – avoid obscuring the scale

Georgia's ASERL libraries: FY12 materials expenditures, compared using different y-axis scales

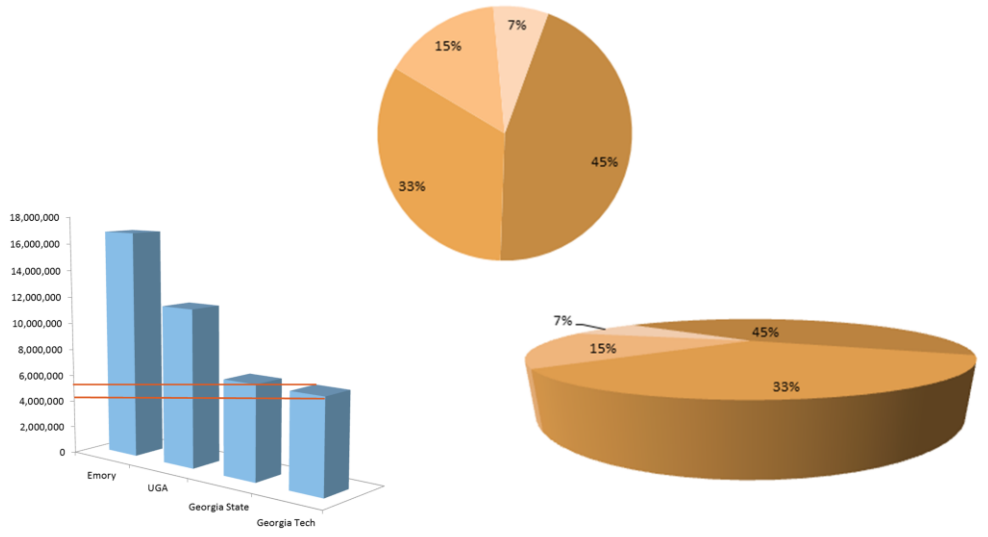


Data source: ASERL Fall 2012 Statistical Index: http://www.aserl.org/about/stats/

On the left, the y-axis scale starts at zero. Emory's expenses are accurately portrayed as about 150% larger than GSU's.

On the right, the y-axis starts at 4,000,0000, which distorts the relationship of Georgia State's expenses in relation to Emory's. In this graph, Emory's expenses look approximately 4.5 times (or 350%) larger.

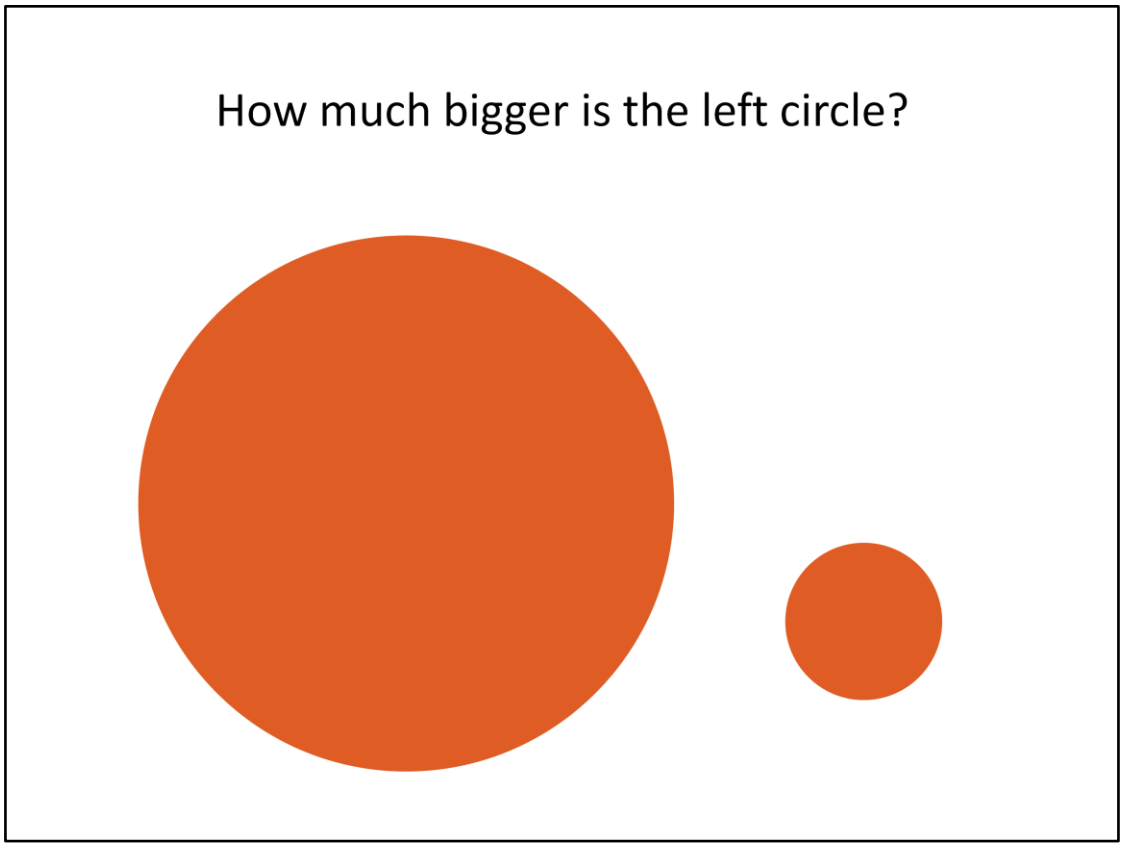
Graphical integrity – avoid 3-D effects



3-D effects are tempting to use to liven up your graphics, but they distort sizes and shapes.

In this 3-D version of the bar chart from the previous slide - where do you measure the top? Do you measure height from the front, back, or middle of the bars?

The flat pie chart shows the proper visual relationship between percentages. I've manipulated the 3-D version so that the second largest segment of 33% now looks the largest. The smallest segment looks even smaller.



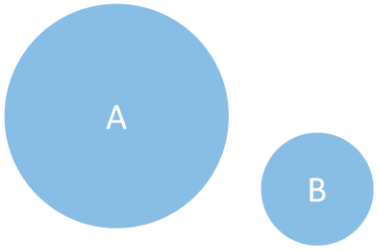
Be aware when you are designing charts with circles - people aren't very good at judging their comparative areas.

The large circle is 12x bigger: an area of 18 compared to an area of 1.5.

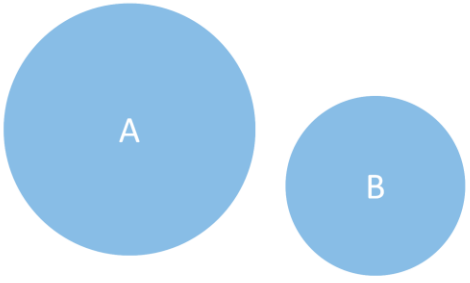
For circles, area = πr^2

Compare areas, not diameters, when drawing circles of relative sizes.
For one circle twice as big as the other:

Incorrect: diameter of circle A is twice the size of circle B.



Correct: area of circle A is twice the size of circle B.



In addition to the difficulties audience members may have in perceiving circle sizes, sometimes comparative circles are incorrectly drawn. Remember to compare the sizes of the areas, not the sizes of the diameters.

On the left, diameters were drawn at 2" and 1". On the right, the areas equal 4" and 2".

Graphical Excellence

Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.

(Tufte, 2001)

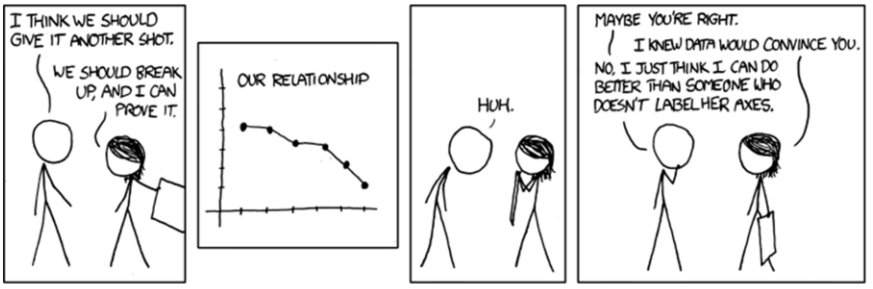
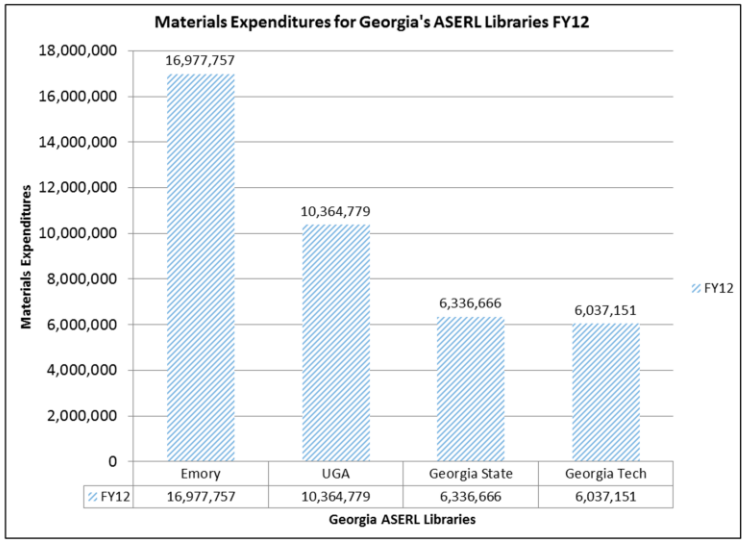


Image source: http://xkcd.com/833/

Presenting your graphics clearly through clean design and coherent labeling is another component of Tufte's graphical excellence.

$$\text{Data-ink ratio} = \frac{\textbf{data-ink}}{\textbf{total ink used to print the graphic}}$$

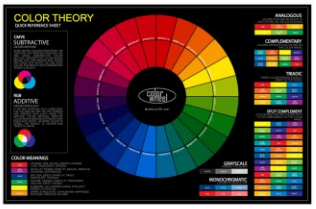


Tufte also coined the terms data-ink, data-ink ratio, and chartjunk.
Data ink is the ink on the graph that represents the data.

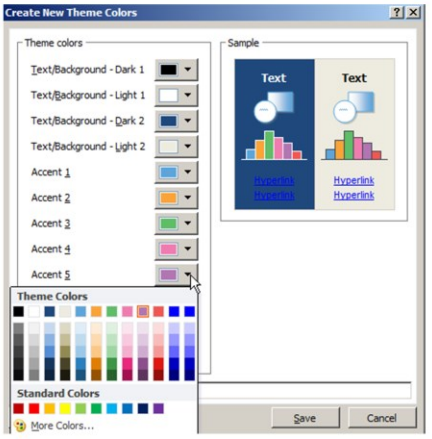
You want to maximize the data-ink ratio and minimize chartjunk. Chartjunk is decoration of the graphics that provides no new information about the data, i.e. grid lines, frames, excess tick marks, keys, vibration/shading/texture effects, redundant chart elements, etc.

Example of what not to do: the ASERL materials expenditures chart has been edited to include many elements that distract the viewer from the data and undermine clear presentation.

Color tips



- Choose a palette and stick with it throughout your presentation
- Don't apply all the colors in your palette at the same time; choose harmonious combinations (shades of the same color, colors on the same side of the color wheel, etc.)
- Use color strategically to signify changes in information or additional levels of data



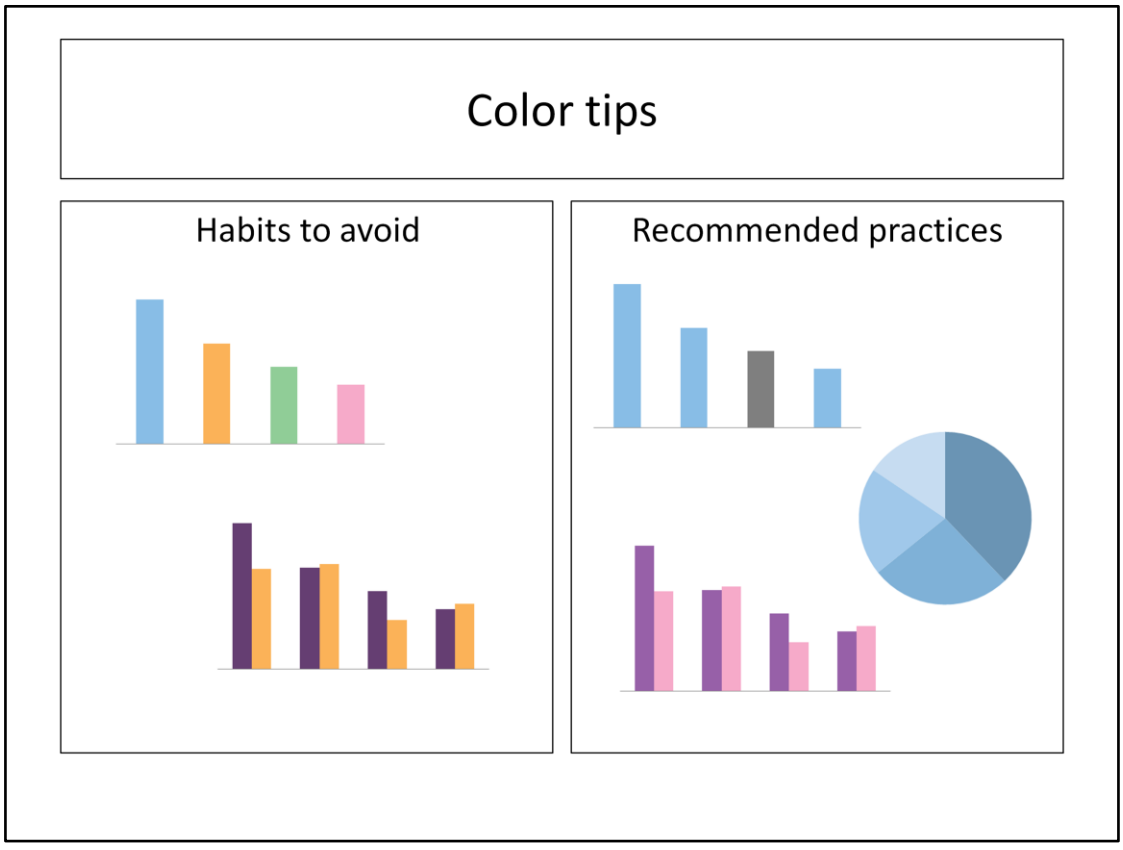
In Excel themes, Accents 1-6 are the chart colors

A color palette for charts should include the basic colors and 3-5 shades of each hue.

S. Few recommends choosing a bright/dark, light, and medium palette of the same hues: use light for shading larger objects (bars, boxes), medium for smaller objects (lines, points), and dark for highlighting, as for a particular bar (p. 344).

Don't choose colors arbitrarily – be strategic; the information you present should determine the colors you choose.

Remember that about 10% of males and 1% of females have some form of color blindness. Most common = red and green both appear brown. Use only one of these, or vary intensities so that differences are distinguishable.

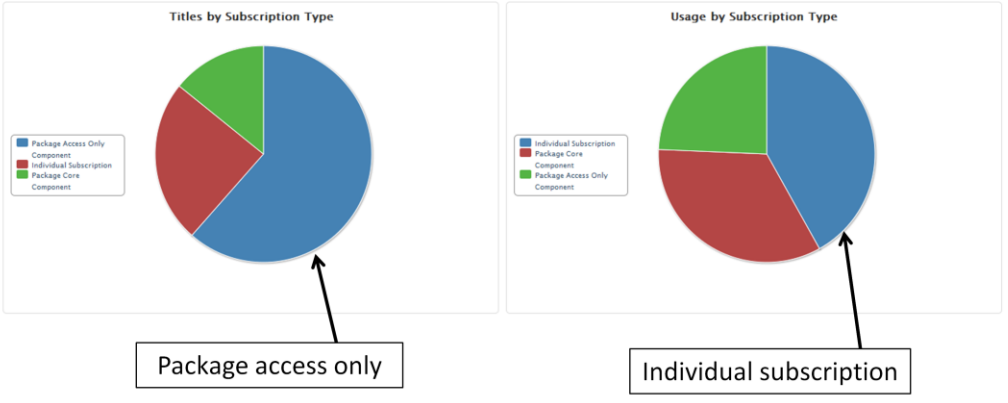


Don’t use multiple colors to represent the same kind of data. Use the same color to represent the same variable – easier to focus on the data.

For a multiple-bar chart, where columns are adjacent, don’t use colors on opposite sides of the color wheel – the contrast is distracting. Use graduating shades of the same color, or colors that are close to each other on the color wheel.

Color tips

Don't use the same color to represent different data entities when displaying charts together.



This example, which grossly abuses color, shows dashboard analytics from one of our vendor software packages. One chart shows titles by subscription type; the other shows usage by subscription type. The journal subscription types for both charts include the same three categories: individual subscriptions, package core, and package access only.

The software automatically orders the pie segments by their values, and the colors are always assigned in the same order. This results in each subscription category being assigned a different color in each chart. This is confusing, particularly since the charts are displayed side by side.

Choosing appropriate display methods

Image source: Dan Piraro's *Bizarro,* (December 1, 2004)

In this section I'll discuss some of the most common chart types and their appropriate uses.

Tables

Tables are the best way to show exact numerical values.

They are often preferable:

- for very small data sets
- when data presentation requires many localized comparisons

No. 854. Domestic Fish and Shellfish Catch and Value by Major Species Caught: 1990 to 2002

Species	Quantity (1,000 lb.)				Value ($1,000)			
	1990	2000	2001	2002	1990	2000	2001	2002
Total	**9,403,571**	**9,068,985**	**9,491,836**	**9,397,154**	**3,521,995**	**3,549,481**	**3,228,285**	**3,092,318**
Fish, total [1]	**8,091,068**	**7,689,661**	**8,242,490**	**8,089,987**	**1,900,097**	**1,594,815**	**1,479,988**	**1,359,392**
Cod: Atlantic	95,881	25,060	33,211	29,841	61,329	26,384	32,086	30,715
Pacific	526,396	530,505	471,711	512,827	91,384	142,330	118,071	96,206
Flounder	254,519	412,723	352,363	372,697	112,921	109,910	105,240	102,370
Halibut	70,454	75,190	77,978	82,044	96,700	143,826	115,169	135,603
Herring, sea; Atlantic	113,095	160,269	209,191	135,871	5,746	9,972	12,717	9,106
Herring, sea; Pacific	108,120	74,835	91,297	78,408	32,178	12,043	13,213	11,534
Menhaden	1,962,160	1,760,498	1,741,430	1,750,609	93,896	112,403	102,690	105,102
Pollock, Alaska	3,108,031	2,606,802	3,179,407	3,341,105	268,344	160,525	230,723	203,696
Salmon	733,146	628,638	722,832	567,179	612,367	270,213	208,926	155,010
Tuna	62,393	50,779	51,854	49,358	105,040	95,176	93,497	84,116
Whiting (Atlantic, silver)	44,500	26,855	28,479	17,622	11,281	11,370	13,232	7,454
Whiting (Pacific, hake)	21,232	452,718	379,304	285,714	1,229	18,809	16,147	13,584
Shellfish, total [1]	**1,312,503**	**1,379,324**	**1,138,512**	**1,178,590**	**1,621,898**	**1,954,666**	**1,711,391**	**1,706,426**
Clams	139,198	118,482	122,764	130,076	130,194	153,973	161,992	167,215
Crabs	499,416	299,006	272,246	307,601	483,837	405,006	381,667	397,695
Lobsters: American	61,017	83,180	73,637	82,252	154,677	301,300	254,334	293,329
Oysters	29,193	41,146	32,673	34,397	93,718	90,667	80,946	89,071
Scallops, sea	39,917	32,747	46,958	53,056	153,696	164,609	175,349	203,707
Shrimp	346,494	332,486	324,481	316,727	491,433	690,453	568,547	460,878
Squid, Pacific	36,082	259,508	191,532	160,677	2,636	27,077	17,834	18,262

[1] Includes other types of fish and shellfish, not shown separately.

Source of Tables 852-854: U.S. National Oceanic and Atmospheric Administration, National Marine Fisheries Service, *Fisheries of the United States*, annual. See also <http://www.st.nmfs.gov/st1/fus/fus01/2002-fus.pdf> (released September 2003).

Image source: U.S. Census Bureau, Statistical Abstract of the United States: 2004-2005

They aren't as visually exciting, but tables are sometimes the most appropriate display method – best way to show exact values.

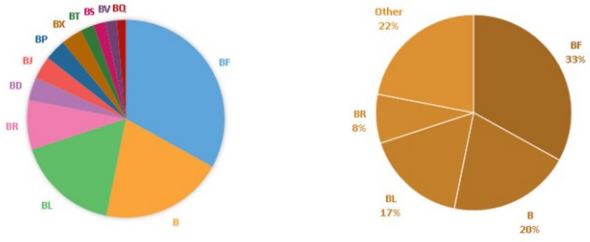
A small table comparing 3 or 4 percentages works as well or better than a pie chart.

One supertable is better than many bar charts if you have large amounts of data where you can make many different comparisons. (Tufte, p. 179)

Pie Charts

- Show percentages
- Should add up to 100%
- Limit to 6 slices
- Put the largest slices in the 12 o'clock position

Circulation distribution of the B call number range in 2012



Use percentages, not values, in your labeling – don't make the user do the math.

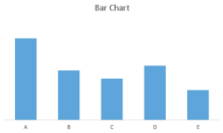
If you have more than 5-6 slices, combine smaller values into an "other" category, as in the chart on the right. If you must display all categories, forgo the pie chart as your display method.

Our eyes gravitate to the top right first. Order slices by size (exception – put the "other" slice last even if it isn't the smallest).

Pie charts are disparaged by many experts, but our audiences are quite familiar with them.

Bar Charts

- Measure discrete quantities
- Usually recommended for comparisons among distinct categories or over time
- Start the scale at zero to avoid distortion

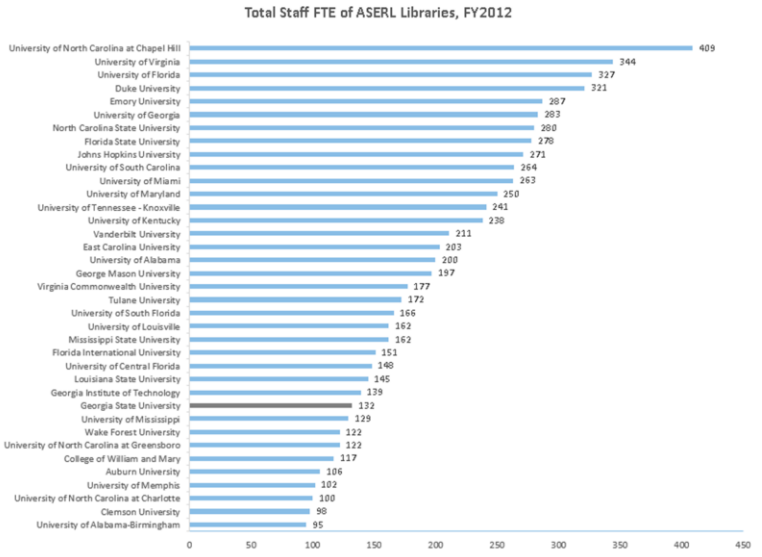


Bar charts – workhorses of quantitative visualization

Some background: 1st bar chart credited to William Playfair (appeared in *The Commercial and Political Atlas,* 1786). The way we view them has changed over time. Up through the early 20th century the area of the bars was the measure of the chart values – widths and heights could both vary. Modern expectation is that the bar length measures the value [widths are uniform and evenly spaced].

Types of bar charts: horizontal, stacked, multi-bar
The latter two are used when your variables have subcategories, e.g. graphing ILL transactions by department, further broken down by patron type.

Horizontal bars work best for long labels



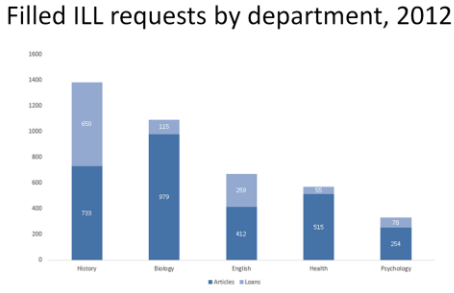
Data source: ASERL Fall 2012 Statistical Index: http://www.aserl.org/about/stats/

If values have long labels, a horizontal bar chart makes your data easier to view – it is hard to read labels vertically or on a slant.

Sort by value before plotting your chart for easier interpretation of how the values relate to each other. Use highlighting to mark important values.

Stacked Bar Charts

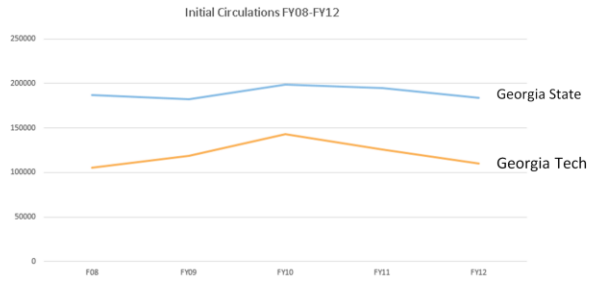
- Use when both the overall total and the individual values of subcategories are important.
- Values in categories at the top of the stacks are hard to compare. Put largest categories at the bottom and label individual values on the chart.



If the overall totals are not critical, a multiple-bar chart is often better than a stacked one.

Line Charts

- Lines are best used to display continuous data series over a period of time
- Well suited for showing trends, acceleration or deceleration, and volatility



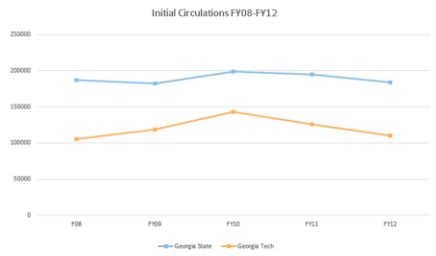
Data source: ASERL Fall 2009-12 Statistical Indexes: http://www.aserl.org/about/stats/

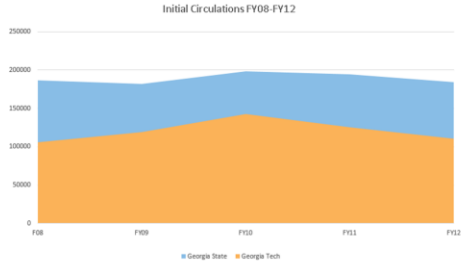
Also called x/y charts – another popular icon of quantitative visual analysis.

These are really just bar charts where the tops of the columns are replaced by lines.

Line Charts

- Avoid marking the dots
- Keep charts to 4 or fewer lines
- Unlike bar charts, the baseline does not have to start at zero (unless you convert it to an area chart, which shows volume)
- Use natural increments like you would for counting (0, 5, 10, 15 not 0, 4, 8, 12, 16)





Data source: ASERL Fall 2009-12 Statistical Indexes: http://www.aserl.org/about/stats

Avoid marking the dots – they are visually distracting.

Keep charts to 4 or fewer lines – use an array of single line charts if you have multiple data series; easier to see the individual lines for comparison.

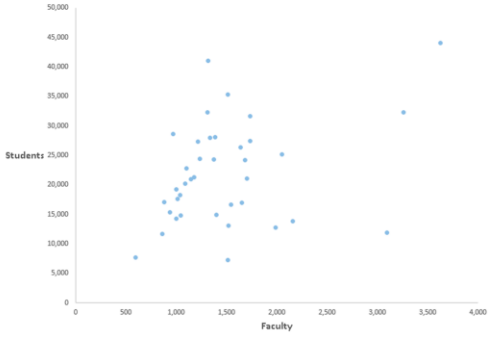
Unlike bar charts, the baseline does not have to start at zero. And many times they shouldn't, as in the case of line charts measuring stock prices. If you plot a range in the thousands starting at zero, small daily changes would be difficult to see. If you convert your line chart to an area chart, however, you do need to start your baseline at zero. In an area chart, you are measuring comparative volumes rather than trend lines.

Line charts are vulnerable to distortion (stretching/flattening) depending on your increment spacing and labeling. Use natural increments like you would for counting: multiples of 5, 10, 100, 1000, etc.

Scatter Plots

- Recommended for investigating and showing correlations
- Strength = their ability to examine more than one variable within same population

Number of full time students vs. faculty at ASERL institutions, FY12

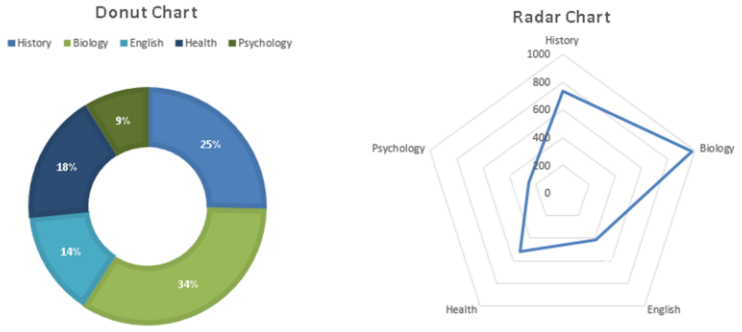


Data source: ASERL Fall 2012 Statistical Index: http://www.aserl.org/about/stats/

One value is plotted on the y axis, the other on the x.

If the points form a band moving from the bottom left to the top right of the chart, it indicates a positive correlation (as one value increases, so does the other). If the points form a band moving from the top left to the bottom right, it indicates a negative correlation (as one value increases, the other decreases).

Few's "silly graphs that are best forsaken"



Filled article requests by department, FY12

This is a teaser slide to entice further reading:
S. Few has a chapter that is delightfully named "silly graphs that are best forsaken." It discusses several charts to avoid, including some that are widely available in popular software packages, such as the donut and the radar/spider chart.

Sources

Few, S. (2012). *Show me the numbers : Designing tables and graphs to enlighten* (2nd ed.). Burlingame, CA: Analytics Press.

Jones, G. E. (1995). *How to lie with charts*. San Francisco: Sybex.

Strange, N. (2007). *Smoke and mirrors : how to bend facts & figures to your advantage*. London, A. & C. Black.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Wong, D. M. (2010). *The Wall Street journal guide to information graphics: The dos and don'ts of presenting data, facts, and figures* (1st ed.). New York: W.W. Norton & Co.

Yau, N. (2011). *Visualize this: The FlowingData guide to design, visualization, and statistics*. Indianapolis, IN: Wiley Pub.